

Cognitively Salient Semantic Relations: Extracting Modifiers of Composite Part Relations

Gerhard Kremer
gerhard.kremer@unitn.it
CIMEC, University of Trento



Introduction

Goal

Providing lexicographers / language learners with salient semantic relation instances for a given concept.

(e.g. dog: tail 😊 vs. heart 😞)

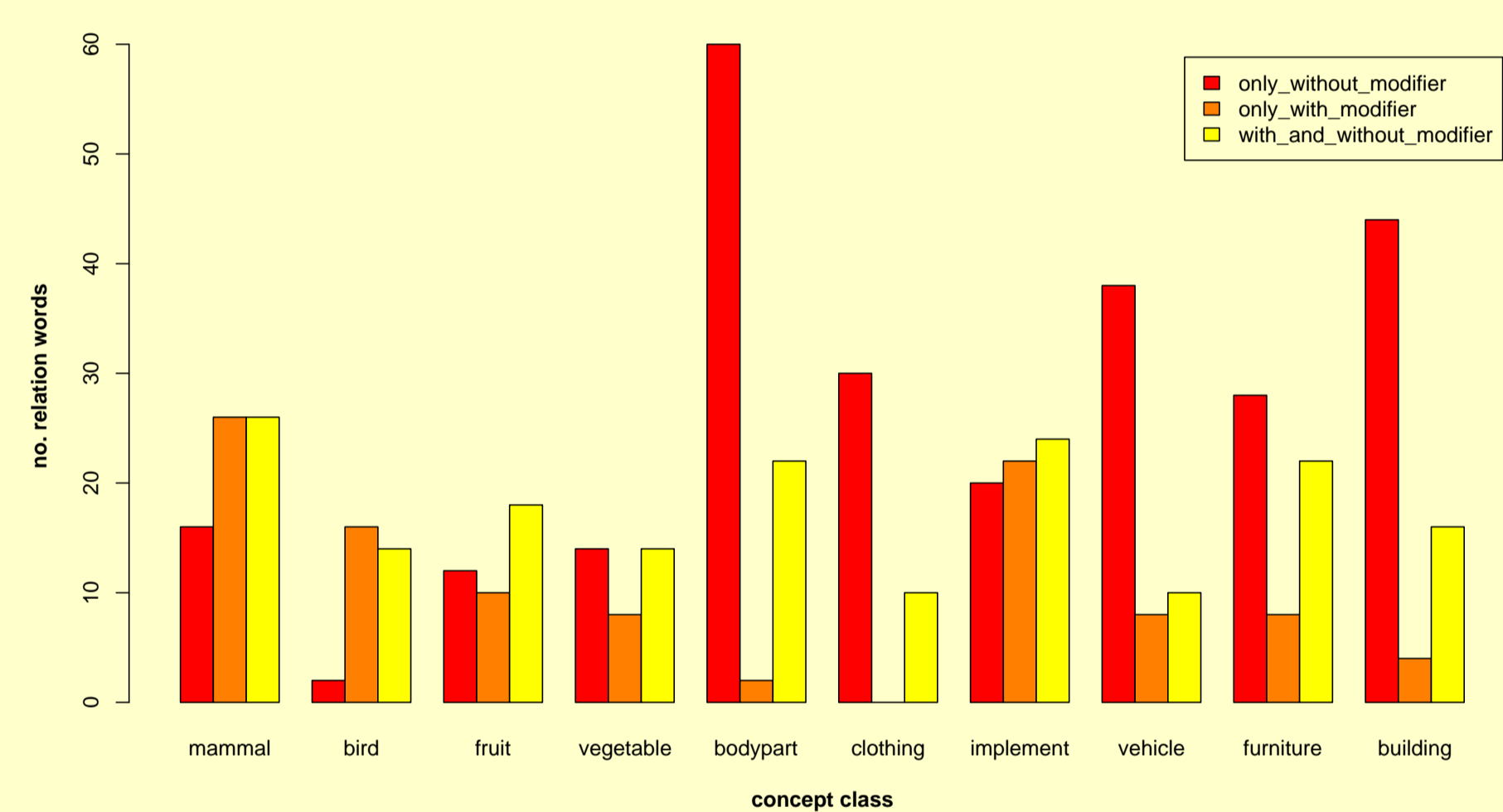
Current Focus

Composite *part* relations: [adj modifier] [noun]

Supporting the Idea

Analysis results of our preceding feature production experiment for DE + IT speakers (collection of descriptions for concrete concepts): Modifiers used further specified a feature – in contrast to multiword expressions.

e.g. dog: wet nose



Further Motivation

Previous automatic extraction algorithms have been concentrating on atomic relation instances – but lexicographic projects make use of composites, e.g. WordNet (≈ 1300 instances), ConceptNet (≈ 68000 instances).

In this poster

Research Question

Assuming that salient part relations have already been identified, which are the most salient modifiers?

Approach Idea

Generate ranked lists of modifiers and select top candidates, using WaCKy web corpus frequencies of modifiers.

Method

Starting from DE data from production experiment –concepts with produced parts– look up corpus frequencies for pairs [adj]..[noun], separated maximally by 3 words:

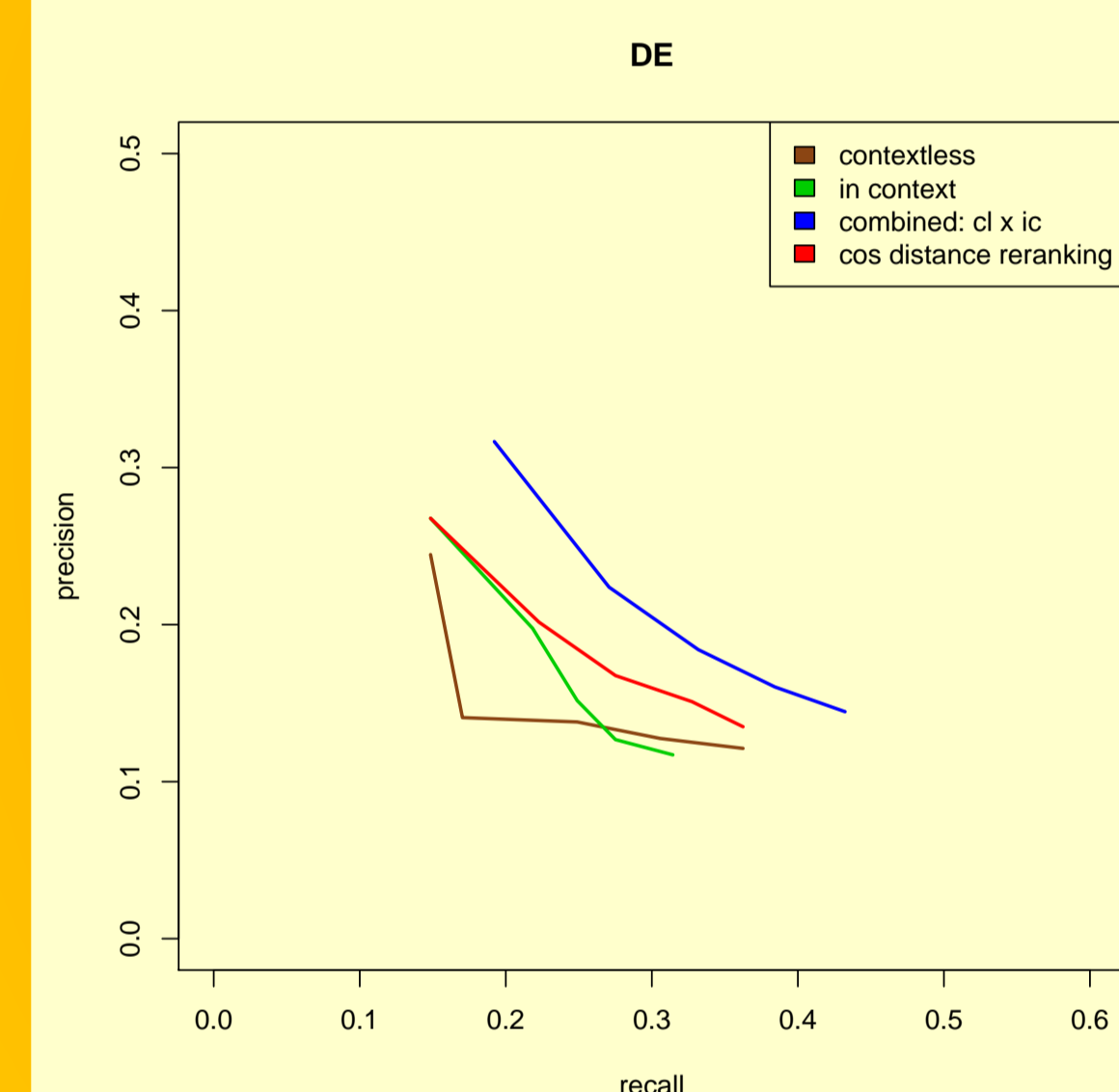
- Within wide context of concept (20 sentences) – concept-specific, but sparsity problem, e.g. bear: white fur, but not broom: long bristles
- Not considering concept context – more exhaustive, but modifiers not concept-specific, e.g. pineapple: thick peel, but owl: yellow eye on low rank

⇒ Combine information: $freq_{combi} = freq_{mod-part\ in\ concept\ context} \times freq_{mod-part\ not\ considering\ concept}$

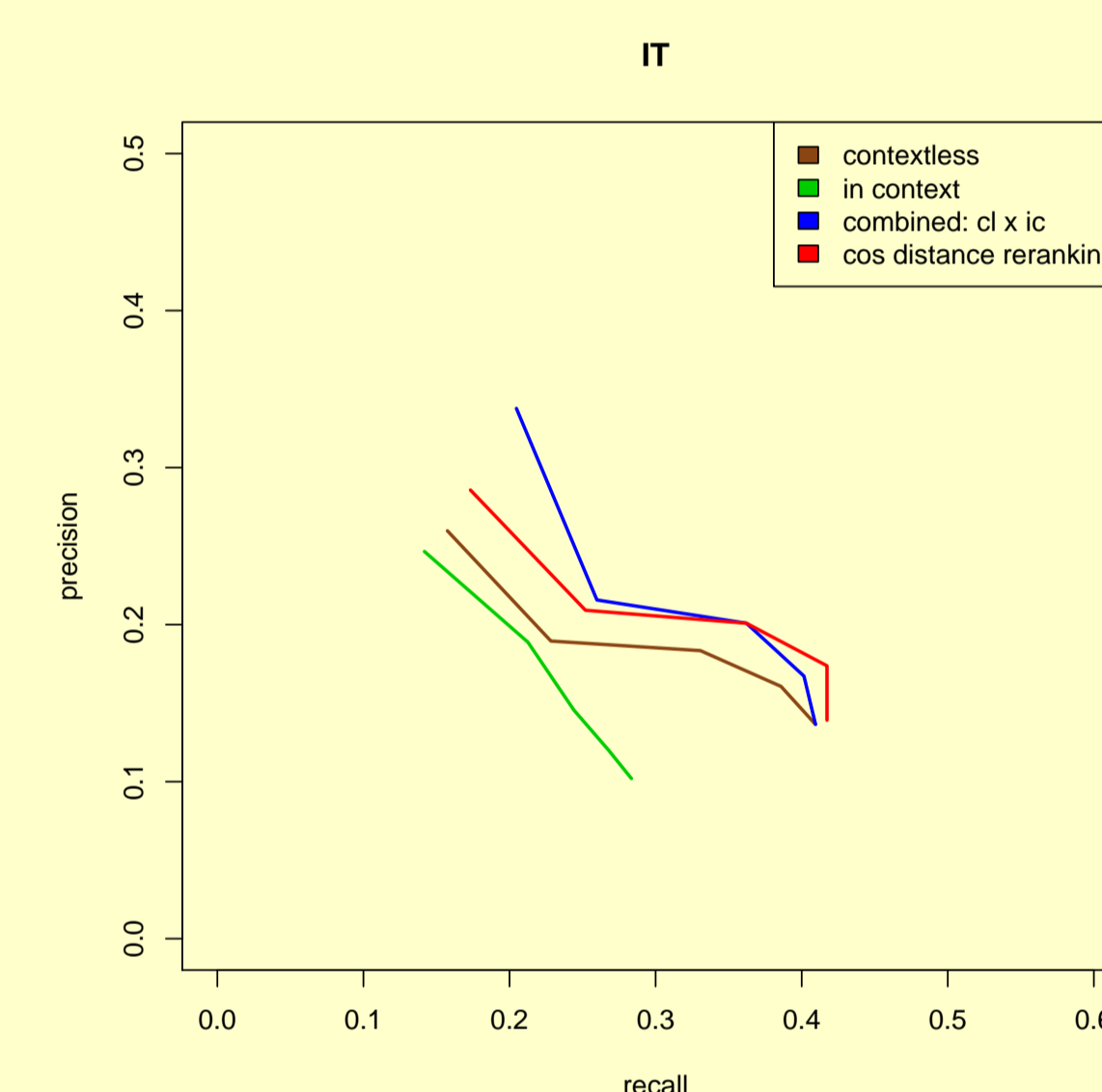
Select top 5 candidates from each of these ranked lists.

Performance

Gold Standard: Composite parts of production experiment (DE)



Method trained on DE data performs similarly well on IT data:

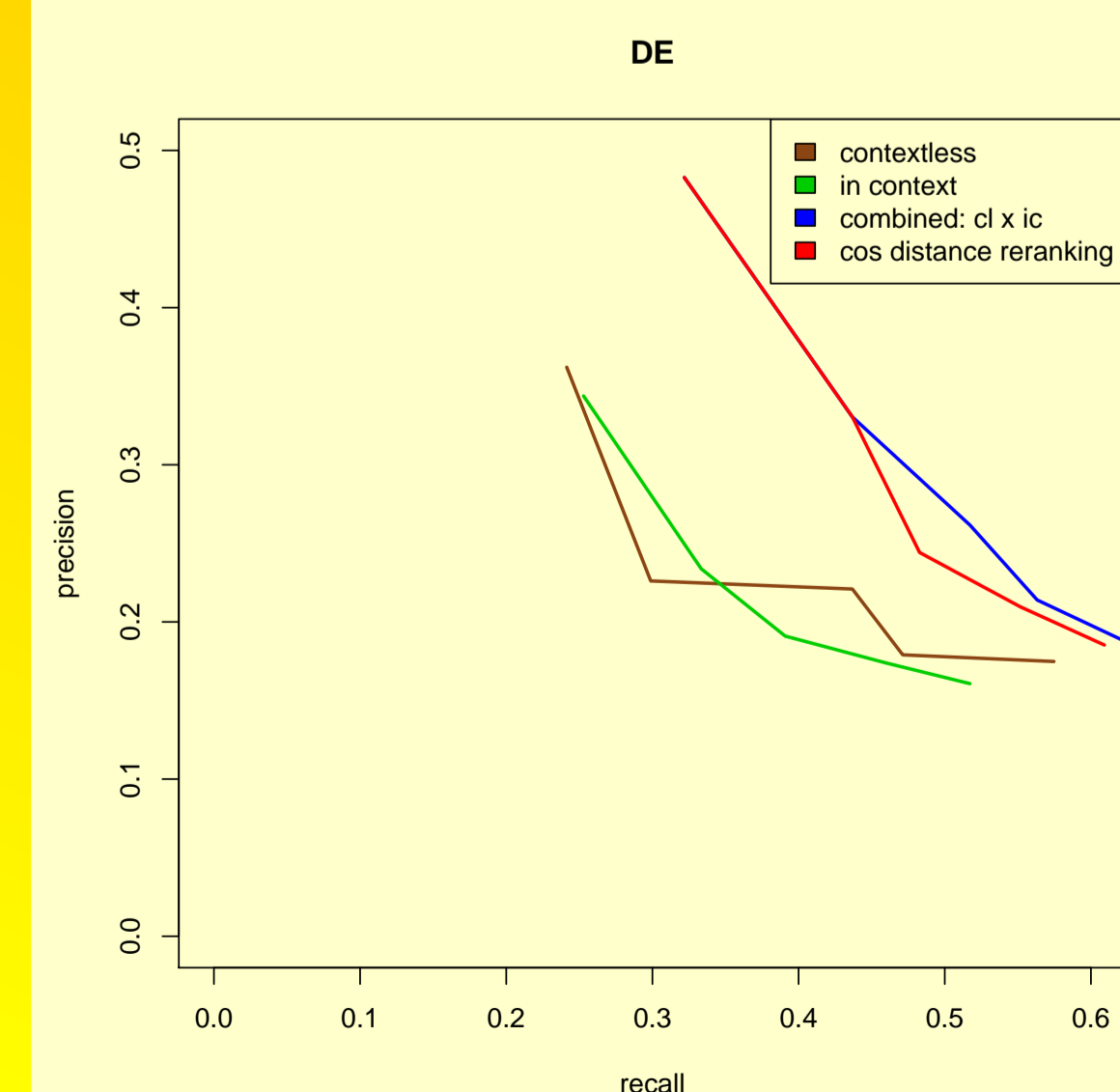


✗ Cosine distance reranking: $freq_{combi} \times (1 + cos-dist)$

Raise combined freq value according to similarities to top 100 contextless modifiers – distance calculation based on number of cooccurrences with nouns.

Performance on New Concepts

Select DE modifiers for the concept–part pairs of McRae et al.'s production experiment.



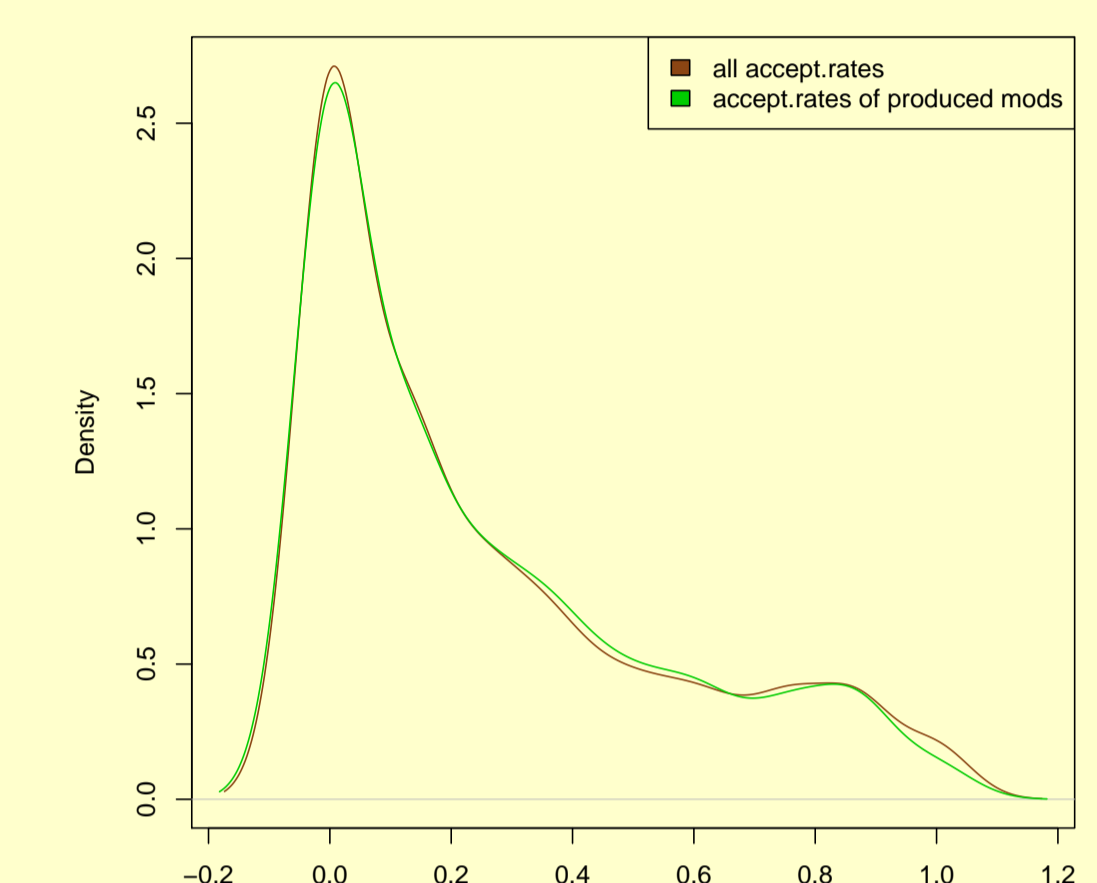
Gold Standard (from judgement experiment):
Modifiers with acceptance rate ≥ 0.75

Production vs. Perception

Acceptance of those DE concept–part–modifier triples that were collected in the production experiment: Small overlap of modifiers that were produced and accepted (46), compared to the numbers of modifiers exclusively produced (53) or accepted (42).

Distribution of acceptance rates only for the triples from the production experiment compared to the overall distribution

T-test: p=0.51



Conclusion

- What is produced quite differs from what is perceived as salient
- Our method yields modifiers of both kinds

Visual Exploration

Only produced:

monkey: short leg, and also bear: white fur

Only accepted:

corn: large leaf, but also bear: white tooth

Next in Line

- Resulting salient modifiers useful for identifying salient parts?
- Adaptability of method to other semantic relation types, and with possible further improvement?

References

- Kremer, G., Abel, A., and Baroni, M. (2008). Cognitively Salient Relations for Multilingual Lexicography. In *Proceedings of the Workshop on Cognitive Aspects of the Lexicon (COGALEX 2008)*, pages 94–101, Manchester, United Kingdom. Coling 2008 Organizing Committee.
- McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic Feature Production Norms for a Large Set of Living and Nonliving Things. *Behaviour Research Methods, Instruments & Computers*, 37(4):547–559.